Introduction to Version Control with GitHub

Nate Wells

Math 243: Stat Learning

September 2nd, 2020

Outline

In today's class, we will. . .

Outline

In today's class, we will...

- Explore the layout of Rstudio

Outline

In today's class, we will...

- Explore the layout of Rstudio
- Discuss version control with GitHub

Outline

In today's class, we will...

- Explore the layout of Rstudio
- Discuss version control with GitHub
- Describe the typical GitHub workflow

Outline

In today's class, we will...

- Explore the layout of Rstudio
- Discuss version control with GitHub
- Describe the typical GitHub workflow
- Practice cloning, pulling, committing, and pushing

Section 1

RStudio

Using RStudio

- RStudio is a convenient interface for interacting with the R programming language

Using RStudio

- RStudio is a convenient interface for interacting with the R programming language
- The easiest way to access RStudio is via the Reed RStudio server:

  https://rstudio.reed.edu

Using RStudio

- RStudio is a convenient interface for interacting with the R programming language
- The easiest way to access RStudio is via the Reed RStudio server:

  https://rstudio.reed.edu

- It is also possible to install R and RStudio locally on your computer.

Using RStudio

- RStudio is a convenient interface for interacting with the R programming language
- The easiest way to access RStudio is via the Reed RStudio server:

  https://rstudio.reed.edu

- It is also possible to install R and RStudio locally on your computer.
  - Instructions: https://www.reed.edu/data-at-reed/resources/R/installr.html

Using RStudio

- RStudio is a convenient interface for interacting with the R programming language
- The easiest way to access RStudio is via the Reed RStudio server:

  https://rstudio.reed.edu

- It is also possible to install R and RStudio locally on your computer.
  - Instructions: https://www.reed.edu/data-at-reed/resources/R/installr.html
- Advantages of RStudio server.

Using RStudio

- RStudio is a convenient interface for interacting with the R programming language
- The easiest way to access RStudio is via the Reed RStudio server:

  https://rstudio.reed.edu

- It is also possible to install R and RStudio locally on your computer.
  - Instructions: https://www.reed.edu/data-at-reed/resources/R/installr.html
- Advantages of RStudio server.
  - Accessible from any computer

Using RStudio

- RStudio is a convenient interface for interacting with the R programming language
- The easiest way to access RStudio is via the Reed RStudio server:

  https://rstudio.reed.edu

- It is also possible to install R and RStudio locally on your computer.
  - Instructions: https://www.reed.edu/data-at-reed/resources/R/installr.html
- Advantages of RStudio server.
  - Accessible from any computer
  - Avoid possibly messy installation

Using RStudio

- RStudio is a convenient interface for interacting with the R programming language
- The easiest way to access RStudio is via the Reed RStudio server:

  https://rstudio.reed.edu

- It is also possible to install R and RStudio locally on your computer.
  - Instructions: https://www.reed.edu/data-at-reed/resources/R/installr.html
- Advantages of RStudio server.
  - Accessible from any computer
  - Avoid possibly messy installation
  - Provides some storage space for files, data, and images

Using RStudio

- RStudio is a convenient interface for interacting with the R programming language
- The easiest way to access RStudio is via the Reed RStudio server:

  https://rstudio.reed.edu

- It is also possible to install R and RStudio locally on your computer.
  - Instructions: https://www.reed.edu/data-at-reed/resources/R/installr.html
- Advantages of RStudio server.
  - Accessible from any computer
  - Avoid possibly messy installation
  - Provides some storage space for files, data, and images
- Advantages of local installation

Using RStudio

- RStudio is a convenient interface for interacting with the R programming language
- The easiest way to access RStudio is via the Reed RStudio server:

  https://rstudio.reed.edu

- It is also possible to install R and RStudio locally on your computer.
  - Instructions: https://www.reed.edu/data-at-reed/resources/R/installr.html
- Advantages of RStudio server.
  - Accessible from any computer
  - Avoid possibly messy installation
  - Provides some storage space for files, data, and images
- Advantages of local installation
  - Much more flexibility and customization

Using RStudio

- RStudio is a convenient interface for interacting with the R programming language
- The easiest way to access RStudio is via the Reed RStudio server:

  https://rstudio.reed.edu

- It is also possible to install R and RStudio locally on your computer.
  - Instructions: https://www.reed.edu/data-at-reed/resources/R/installr.html
- Advantages of RStudio server.
  - Accessible from any computer
  - Avoid possibly messy installation
  - Provides some storage space for files, data, and images
- Advantages of local installation
  - Much more flexibility and customization
  - Can be used after you graduate from Reed

RMarkdown

- RStudio can also create and edit a special document type called **RMarkdown files** (.Rmd) which can seemlessly integrate code, results, and prose

RMarkdown

- RStudio can also create and edit a special document type called **RMarkdown files** (.Rmd) which can seemlessly integrate code, results, and prose
- These features allow efficient generation of *reproducible* and *redistributable* results

RMarkdown

- RStudio can also create and edit a special document type called **RMarkdown files** (.Rmd) which can seemlessly integrate code, results, and prose
- These features allow efficient generation of *reproducible* and *redistributable* results
- RMarkdown can be used to create reports, assignments, journal articles, books, and presentations (like this one!)

RMarkdown

- RStudio can also create and edit a special document type called **RMarkdown files** (.Rmd) which can seemlessly integrate code, results, and prose
- These features allow efficient generation of *reproducible* and *redistributable* results
- RMarkdown can be used to create reports, assignments, journal articles, books, and presentations (like this one!)
- RMarkdown can output a variety of file types: .html, .pdf, .doc, and more

New RMarkdown Files

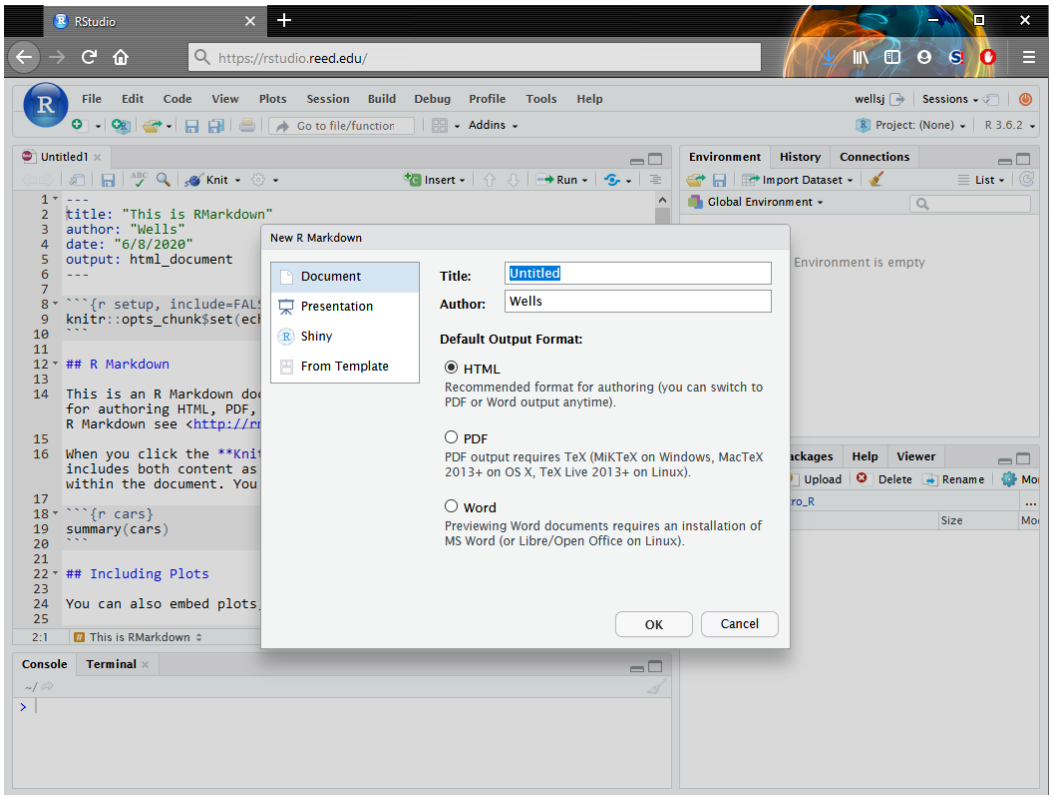
- Open a new .Rmd file by selecting **File** -> **New File** -> **R Markdown** from the navigation bar at the top of the screen

New RMarkdown Files

- Open a new .Rmd file by selecting **File** -> **New File** -> **R Markdown** from the navigation bar at the top of the screen
- You will be presented with a variety of options for output format and document type, but to begin, select "Document" and "pdf" output

New RMarkdown Files

- Open a new .Rmd file by selecting **File** -> **New File** -> **R Markdown** from the navigation bar at the top of the screen
- You will be presented with a variety of options for output format and document type, but to begin, select "Document" and "pdf" output



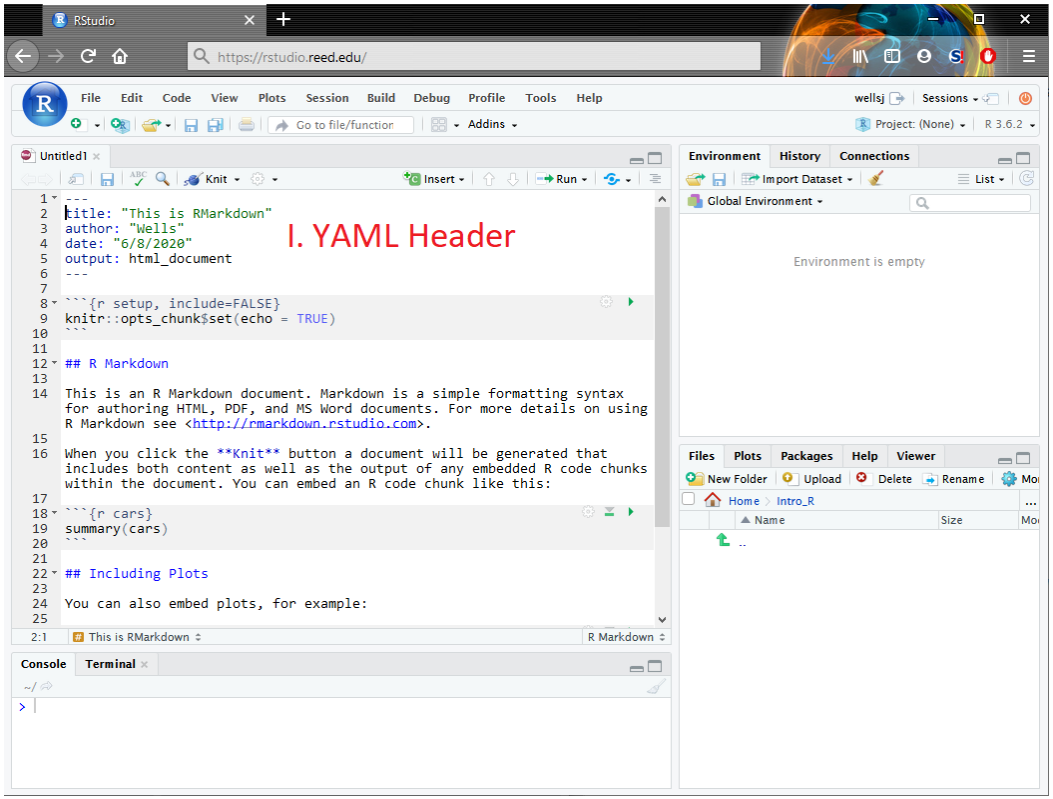
Anatomy of RMarkdown

The standard .Rmd file contains three types of content

Anatomy of RMarkdown

The standard .Rmd file contains three types of content

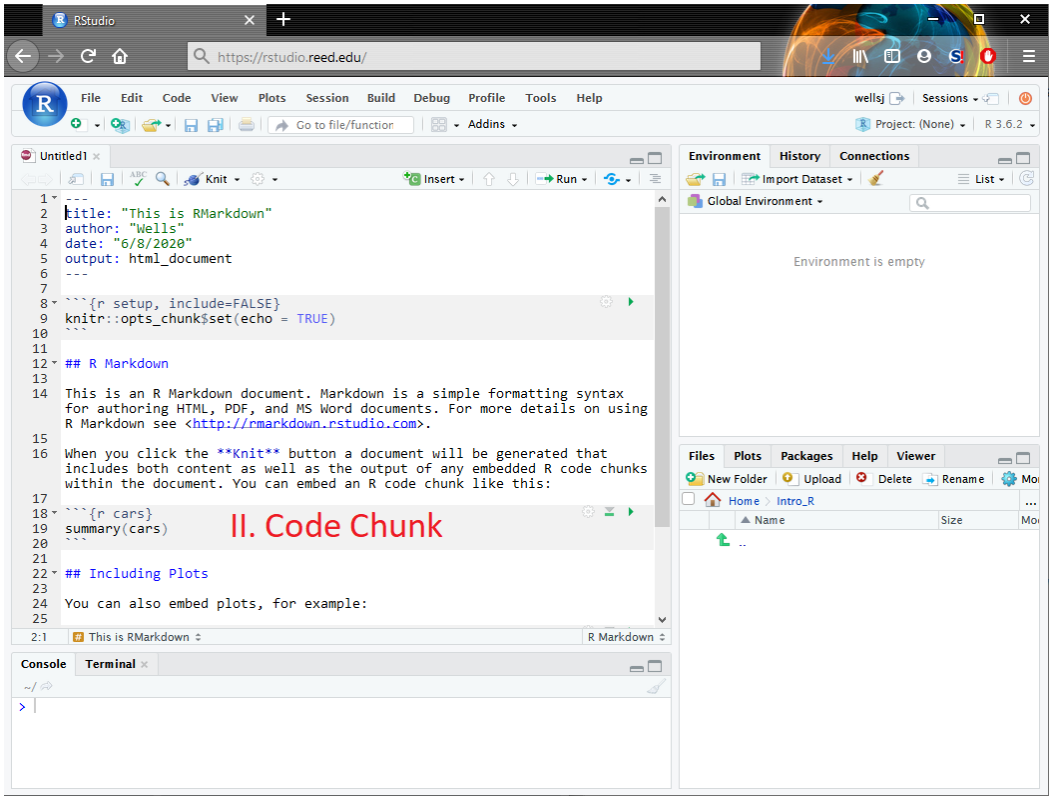
1. A YAML header surrounded by ---



Anatomy of RMarkdown

The standard .Rmd file contains three types of content

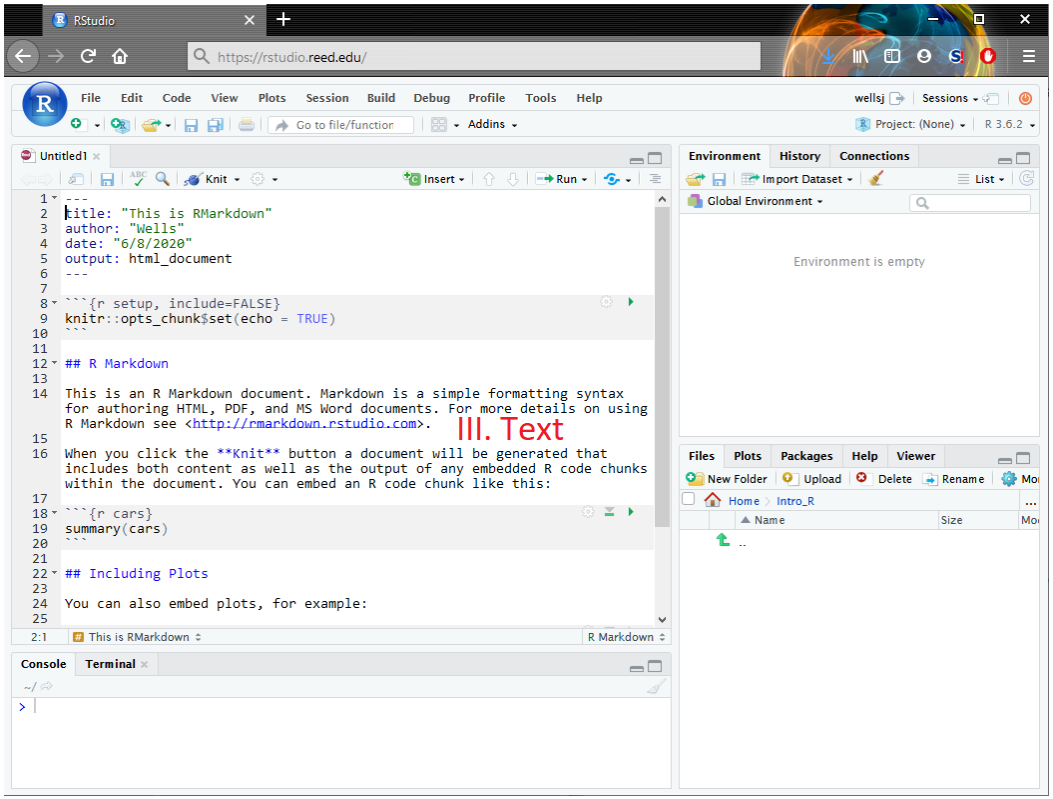
2. **Chunks** of R code surrounded by ```



Anatomy of RMarkdown

The standard .Rmd file contains three types of content

3. Text formatted with simple markdown syntax like `*italics*` or `**bold**`

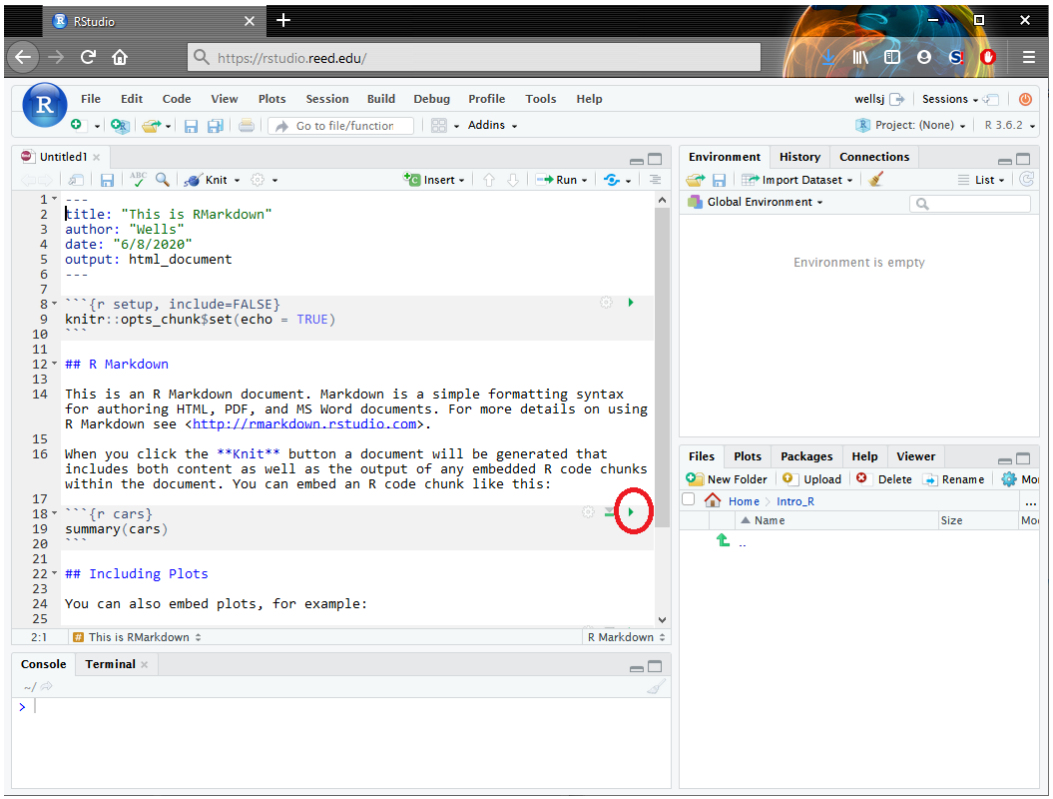


Using RMarkdown

- Clicking the little green arrow in the upper right corner of a code chunk sends the code to the console (just like selecting "run" on an R script)

Using RMarkdown

- Clicking the little green arrow in the upper right corner of a code chunk sends the code to the console (just like selecting "run" on an R script)

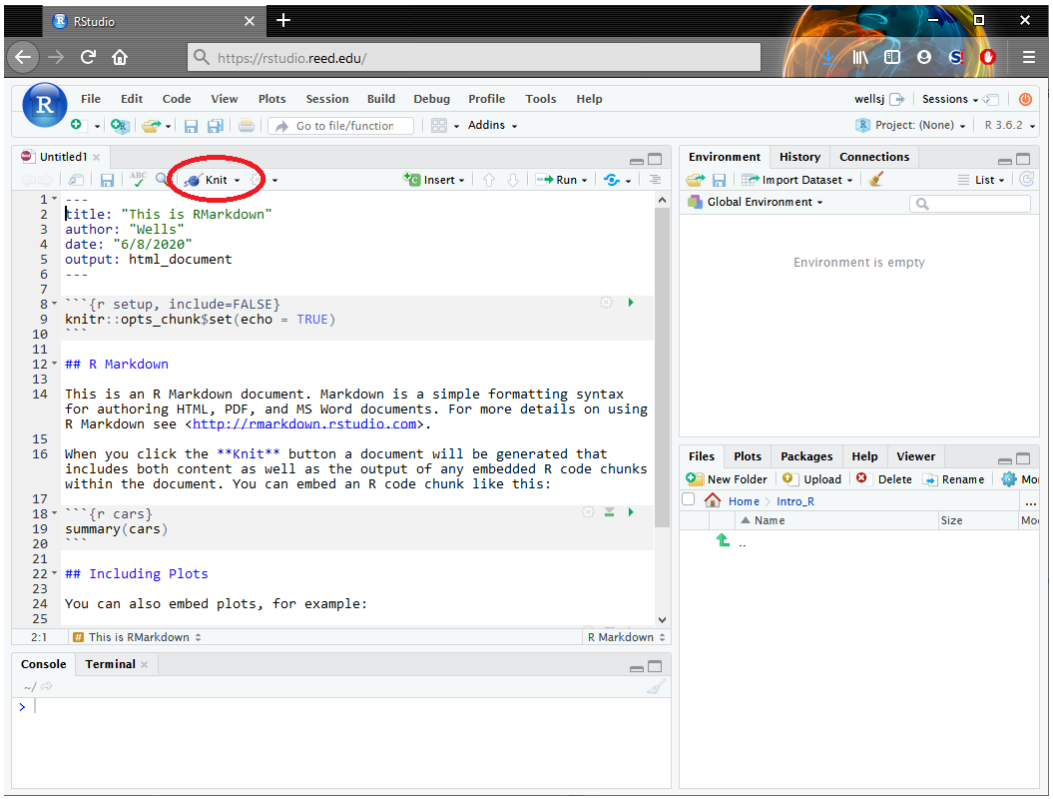


Using RMarkdown

- Alternatively, to generate the output document, select "Knit" at the top of the Editor pane (This is similar to building a .pdf file using LaTeX)

Using RMarkdown

- Alternatively, to generate the output document, select "Knit" at the top of the Editor pane (This is similar to building a .pdf file using LaTeX)



Section 2

Version Control

What is version control?

- Git is one of the most popular version control systems, used by programmers, statisticians, data scientists and more!

What is version control?

- Git is one of the most popular version control systems, used by programmers, statisticians, data scientists and more!
- Git tracks the evolution of a document (and its file system) through a series snapshots (called **commits**)

What is version control?

- Git is one of the most popular version control systems, used by programmers, statisticians, data scientists and more!
- Git tracks the evolution of a document (and its file system) through a series snapshots (called **commits**)
- These commits make it easier to save and compare different versions of a document, as well as to restore a previous version

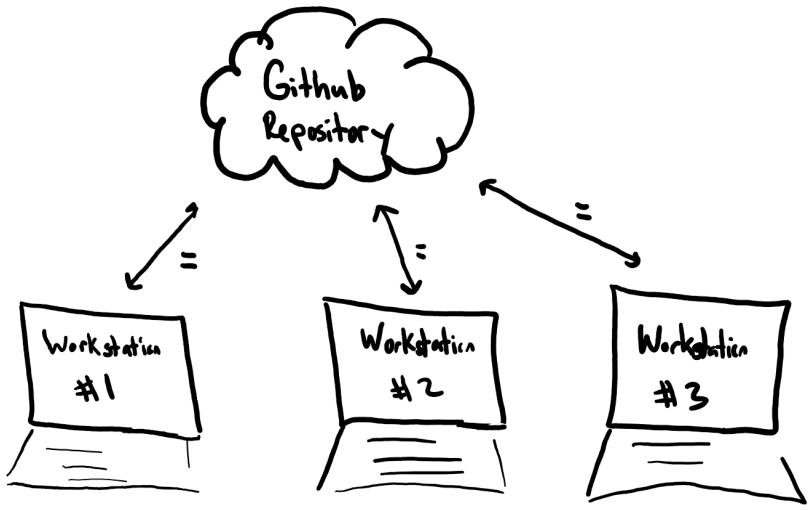
What is version control?

- Git is one of the most popular version control systems, used by programmers, statisticians, data scientists and more!
- Git tracks the evolution of a document (and its file system) through a series snapshots (called **commits**)
- These commits make it easier to save and compare different versions of a document, as well as to restore a previous version
- Collaborators work independently on their version of the document, and then sync regularly to a common version

Why version control?

There two main reasons you may want to use version control:

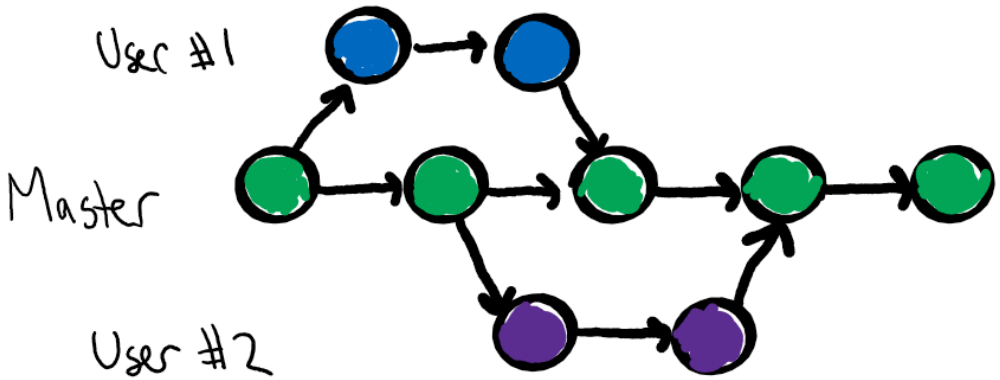
1. **Individual Use**. You have a document you will iterate on frequently, that references several other images and/or data sets, or that you want available on multiple devices



Why version control?

There two main reasons you may want to use version control:

2. **Group Use**. You are collaborating with several other people on a common document, need to synthesize several versions of the document, and comment on changes.



How to version control?

- We will use GitHub (an online interface and storage space for using git) for version control

How to version control?

- We will use GitHub (an online interface and storage space for using git) for version control
  - Think of GitHub as a shared google drive

How to version control?

- We will use GitHub (an online interface and storage space for using git) for version control
  - Think of GitHub as a shared google drive
- Files are stored on GitHub in **respositories** (a nested set of folders)

How to version control?

- We will use GitHub (an online interface and storage space for using git) for version control
  - Think of GitHub as a shared google drive
- Files are stored on GitHub in **respositories** (a nested set of folders)
- Users can upload/download to GitHub using the web interface

How to version control?

- We will use GitHub (an online interface and storage space for using git) for version control
  - Think of GitHub as a shared google drive
- Files are stored on GitHub in **respositories** (a nested set of folders)
- Users can upload/download to GitHub using the web interface
- Or they can interact with GitHub via a client, like RStudio or GitHub Desktop

GitHub Workflow

1. Users **clone** the respositories from GitHub to their computer (or to the RStudio server).

GitHub Workflow

1. Users **clone** the respositories from GitHub to their computer (or to the RStudio server).
2. If the user has already cloned the repo previously, they instead **pull** any commits from the repo to their machine.

GitHub Workflow

1. Users **clone** the respositories from GitHub to their computer (or to the RStudio server).
2. If the user has already cloned the repo previously, they instead **pull** any commits from the repo to their machine.
3. Then, users revise their documents as usual.

GitHub Workflow

1. Users **clone** the respositories from GitHub to their computer (or to the RStudio server).
2. If the user has already cloned the repo previously, they instead **pull** any commits from the repo to their machine.
3. Then, users revise their documents as usual.
4. Occasionally, the user makes a **commit** of their work.

GitHub Workflow

1. Users **clone** the respositories from GitHub to their computer (or to the RStudio server).
2. If the user has already cloned the repo previously, they instead **pull** any commits from the repo to their machine.
3. Then, users revise their documents as usual.
4. Occasionally, the user makes a **commit** of their work.
5. After making several commits (and especially when done working for the session), the user will **push** the commits to the GitHub repository.

Section 3

Practice

Working Solo

Follow the instructions on the Working Solo document found under today's class (Wed 9-1) on the schedule page of the course website.

Section 4

Homework

GitHub Classroom

We will use GitHub Classroom to distribute and submit homework for this course

GitHub Classroom

We will use GitHub Classroom to distribute and submit homework for this course

1. To obtain each homework assignment, use that assignment's link posted in the #annoucements channel on Slack (since anyone with the link can create an assignment, the link won't be posted on the course website)

GitHub Classroom

We will use GitHub Classroom to distribute and submit homework for this course

1. To obtain each homework assignment, use that assignment's link posted in the #annoucements channel on Slack (since anyone with the link can create an assignment, the link won't be posted on the course website)
2. Follow the instructions on the page to create your own repo for that assignment in our Math 243 organization.

GitHub Classroom

We will use GitHub Classroom to distribute and submit homework for this course

1. To obtain each homework assignment, use that assignment's link posted in the #annoucements channel on Slack (since anyone with the link can create an assignment, the link won't be posted on the course website)
2. Follow the instructions on the page to create your own repo for that assignment in our Math 243 organization.
3. Clone the repo either to the RStudio server or your computer (following similar steps to the *Working Solo* activity)

GitHub Classroom

We will use GitHub Classroom to distribute and submit homework for this course

1. To obtain each homework assignment, use that assignment's link posted in the #annoucements channel on Slack (since anyone with the link can create an assignment, the link won't be posted on the course website)
2. Follow the instructions on the page to create your own repo for that assignment in our Math 243 organization.
3. Clone the repo either to the RStudio server or your computer (following similar steps to the *Working Solo* activity)
4. Work on the assignment on the server or on your computer, making frequent commits.

GitHub Classroom

We will use GitHub Classroom to distribute and submit homework for this course

1. To obtain each homework assignment, use that assignment's link posted in the #annoucements channel on Slack (since anyone with the link can create an assignment, the link won't be posted on the course website)
2. Follow the instructions on the page to create your own repo for that assignment in our Math 243 organization.
3. Clone the repo either to the RStudio server or your computer (following similar steps to the *Working Solo* activity)
4. Work on the assignment on the server or on your computer, making frequent commits.
5. When you are finished working, push your commits to repo on GitHub organization. Congrats, you've submitted your assignment! (Note: if you don't push, your work will not be available to the grader)

GitHub Classroom

We will use GitHub Classroom to distribute and submit homework for this course

1. To obtain each homework assignment, use that assignment's link posted in the #annoucements channel on Slack (since anyone with the link can create an assignment, the link won't be posted on the course website)
2. Follow the instructions on the page to create your own repo for that assignment in our Math 243 organization.
3. Clone the repo either to the RStudio server or your computer (following similar steps to the *Working Solo* activity)
4. Work on the assignment on the server or on your computer, making frequent commits.
5. When you are finished working, push your commits to repo on GitHub organization. Congrats, you've submitted your assignment! (Note: if you don't push, your work will not be available to the grader)
6. You will recieve feedback on the assignment via Pull Request (to be discussed later)